Introducción a la Teoría Ergódica

Matilde Martínez

Universidad de la República, Uruguay

Encuentro de Jóvenes Topólogos 2019

1. Introducción

La *dinámica* es el estudio de los sistemas que evolucionan en el tiempo. El tiempo puede ser discreto o continuo, pasado o futuro. Para su estudio matemático, los sistemas dinámicos se suelen dividir en dos grandes clases: *discretos* y *continuos*.

Cuando hablamos de un sistema dinámico discreto, nos referimos a una acción del grupo $\mathbb{Z}$ o del semigrupo $\mathbb{N}$ en un espacio X. Actuar con 1 da una función $T : X \to X$, y el estudio de la dinámica es el estudio de los *iterados* de T, o las composiciones de T consigo misma. Así, si $x \in X$ y n es un número natural o entero,

$$n \cdot x = T^n(x)$$

donde $T^0 = id_X$, $T^n = T \circ \cdots \circ T$ si $n > 0$ y $T^n = T^{-1} \circ \cdots \circ T^{-1}$ si $n < 0$.

Un sistema dinámico continuo es una acción de $\mathbb{R}$ en X, es decir, un flujo. Los flujos surgen naturalmente, por ejemplo, como soluciones a ecuaciones diferenciales ordinarias autónomas $\dot{x} = F(x)$.

Más modernamene, y por extensión de estos ejemplos, suele llamarse *sistema dinámico* a cualquier acción de un grupo, semigrupo o pseudogrupo.

En este curso consideraremos un sistema dinámico discreto dado por una transformación $T : X \to X$. Puede provenir del estudio de un sistema físico cuyo *espacio de estados* es X. Si en el instante inicial el sistema se encuentra en el estado x, un instante después se hallará en el estado $T(x)$, luego en $T^2(x)$, y así sucesivamente. Para entender la evolución del sistema, estudiamos las *órbitas* de T. Si $x \in X$, la órbita futura de x es

$$o_+(x) = \{T^n(x) : \ n \geq 0\}.$$

Cuando T es invertible, se puede definir análogamente la órbita pasada $o_-(x)$, y la órbita $o(x)$ que es unión de ambas. Un punto es fijo si es el único elemento de su órbita, y periódico si su órbita es finita.

En el estudio de la dinámica importan las *propiedades asintóticas* de las órbitas de la acción, o sea, las propiedades "que se observan a tiempos largos". Por ejemplo, si X es un espacio topológico, podemos preguntarnos dónde acumula $o(x)$. Un punto es recurrente si su órbita pasa infinitas veces arbitrariamente cerca de él. Si X es métrico podemos decir que x es estable si las órbitas de puntos cercanos a x se mantienen uniformemente cerca de los iterados de x. Son todas estas propiedades dinámicas interesantes.

La *teoría ergódica* es el estudio de los sistemas dinámicos desde la perspectiva de la teoría de la probabilidad. El sistema dado por $T : X \to X$ es determinista, pero hay factores como la falta de información, la incapacidad de hacer mediciones precisas o la complejidad del sistema que hacen que sea útil estudiarlo probabilísticamente.

Este enfoque nació en la segunda mitad del siglo XIX, con la mecánica estadística. Hasta ese momento, la teoría cinética de los gases intentaba describir la termodinámica en términos de la mecánica. A partir de ella, Ludwig Boltzmann desarrolló la mecánica estadística, que usa la probabilidad para estudiar sistemas mecánicos cuyo estado exacto es incierto. La famosa *hipóteis de Botzmann*, formulada en 1868, se refiere a un sistema conservativo con una cantidad enorme de partículas, tal como un gas, y describe la distribución de los momentos de estas partículas. Así, en estos sistemas extremadamente complejos, se renuncia a la pretensión de entender el comportamiento de cada partícula individualmente, y se apunta a entender propiedades globales del sistema expresadas en el lenguaje de las probabilidades.

La teoría ergódica propiamente dicha nació en 1890, cuando Henri Ponicaré probó, estudiando la mecánica celeste, el primer y más básico resultado de la misma: el Teorema de Recurrencia de Poincaré. No fue sino hasta 50 años después que George D. Birkhoff y John von Neumann establecieron los *teoremas ergódicos* que llevan sus nombres. En 1940, Claude E. Shannon escribió un artículo intitulado *A mathematical theory of communication*, trabajo fundacional de la teoría de la información. Basado en las ideas de Shannon, en 1959, Andréi Kolmogórov definió el concepto matemático de entropía, que es una cantidad que mide la complejidad de un sistema. El estudio del mismo fue profundizado en los años subsiguientes por el mismo Kolmogórov y su alumno Yákov Sinai. En esa época, la década de los 1960s, hizo eclosión la moderna teoría de los sistemas dinámicos, con su énfasis en el estudio de la dinámica *caótica*. En estos sistemas, donde el comportamiento de las órbitas individuales es altamente impredecible, la teoría ergódica encontró su objeto de estudio natural. Se desarrolló intensamente, por ejemplo en el estudio de los billares a partir de los años 1980s.

Actualmente, la teoría ergódica es un activo campo de estudio que interactúa con otras áreas de la matemática de formas a veces sorprendentes. Para ejemplificar esto, citaremos dos teoremas de enunciado elemental cuyas demostraciones son “ergódicas”:

Teorema (Green y Tao, 2004). *Hay progresiones aritméticas arbitrariamente largas de números primos.*

Teorema (Conjeturado por Oppenheim en 1929, demostrado por Margulis en 1987). *Sea Q una forma cuadrática irracional indefinida en R^3 (como por ejemplo $Q(n_1, n_2, n_3) = n_1^2 + n_2^2 - \sqrt{2}n_3^2$). Entonces*

$$\{Q(n_1, n_2, n_3) : \ n_1, n_2, n_3 \in \mathbb{N}\}$$

es denso en $\mathbb{R}$.

2. Introducción a la dinámica medible

2.1. Dinámica de los automorfismos de medida

Consideremos un conjunto X y una σ-álgebra $\mathcal{A}$ de subconjuntos de X. En el espacio medible $(X, \mathcal{A})$, tomemos una transformación medible $T : X \to X$.

Recordemos que una medida de probabilidad en $(X, \mathcal{A})$ es una función $\mu : \mathcal{A} \to [0, 1]$ que cumple las dos propiedades siguientes:

- $\mu(X) = 1$.
- Si $\{A_n\}$ es una sucesión de conjuntos dos a dos disjuntos, $\mu(\cup_n A_n) = \sum_n \mu(A_n)$.

Cuando hablemos de una medida, siempre nos estaremos refiriendo a una medida de probabilidad.

Definición 1. Si μ es una medida de probabilidad en $(X, \mathcal{A})$, decimos que "μ es invariante por T", que "T preserva a μ", que "μ es T-invariante" o que "T es un automorfismo de $(X, \mathcal{A}, \mu)$" si para todo conjunto medible A

$$\mu(T^{-1}(A)) = \mu(A).$$

Observemos que si T es invertible, lo anterior implica que $\forall A \in \mathcal{A}$, $\mu(T(A)) = \mu(A)$.

Ejemplo 1. Tomemos $T : [0, 1] \to [0, 1]$ dada por

$$T(x) = \begin{cases} x/2 & \text{si} \quad x \neq 0 \\ 1 & \text{si} \quad x = 0. \end{cases}$$

Esta transformación es medible con respecto a la σ-álgebra de Borel de $[0, 1]$, y no tiene medidas invariantes. Para verlo, escribamos al intervalo $[0, 1]$ como la unión disjunta $[0, 1] = \{0\} \cup (\cup_{n \geq 0} A_n)$, donde $A_n = (\frac{1}{2^{n+1}}, \frac{1}{2^n}]$ para $n \geq 0$. Como para cada n se tiene que $T^{-1}(A_{n+1}) = A_n$, si una medida de probabilidad μ es invariante por T debe cumplirse que $\mu(A_n) = 0\ \forall n \geq 0$. Por lo tanto, $\mu(\{0\}) = 1$. Pero si la medida es invariante, $1 = \mu(\{0\}) = \mu(T^{-1}(\{0\})) = \mu(\emptyset) = 0$, lo cual es absurdo.

Sin embargo, para una amplia clase de transformaciones medibles hay medidas invariantes. Concretamente:

Teorema. *Sean X un espacio métrico compacto y $T : X \to X$ continua. Entonces existen medidas de probabilidad T-invariantes en la σ-álgebra de Borel de X.*

Ejemplo 2. En la circunferencia $S^1 = \mathbb{R}/\mathbb{Z}$ consideremos la transformación dada por $T(x) = x + a$, donde a es un número fijo. Si pensamos en la circunferencia como $S^1 = \{z \in \mathbb{C} : \ |z| = 1\}$, T es la rotación de ángulo $2\pi a$ dada por $T(e^{2\pi ix}) = e^{2\pi i(x+a)}$. La medida de Lebesgue en la circunferencia es T-invariante. Cuando a es irracional, se puede ver que es la única medida invariante. Sin embargo, cuando $a = p/q$ es racional, todas las órbitas de T son finitas, de exactamente q elementos si la fracción p/q es irreducible. Cada órbita soporta una medida invariante, que da masa $1/q$ a cada uno de sus puntos.

Para esta transformación como para cualquier otra, una combinación lineal convexa, finita o infinita, de medidas invariantes da una nueva medida invariante.

Definición 2. Sean $(X, \mathcal{A})$ un espacio medible, $T : X \to X$ una transformación medible y μ una medida de probabilidad T-invariante. Decimos que $A \in \mathcal{A}$ es invariante si $T^{-1}(A) = A$ a menos de un conjunto de medida cero; es decir, si la diferencia simétrica entre A y $T^{-1}(A)$ tiene medida μ igual a cero. Decimos que μ es ergódica si todo conjunto medible A que es T-invariante tiene medida 0 o 1.

Un conjunto T-invariante es un conjunto saturado por órbitas. Lo que la definición de *ergodicidad* nos dice es que X no puede descomponerse como la unión de dos conjuntos disjuntos de medida positiva donde T pueda estudiarse por separado. Es una condición sobre la medida pues distintas medidas "ven" distintos conjuntos del espacio. Las medidas ergódicas son suficientes para entender la dinámica desde el punto de vista medible, pues se puede probar que cualquier medida invariante es combinación lineal convexa (posiblemente infinita, dada por una integral) de medidas ergódicas. La afirmación precisa es que las medidas invariantes forman un conjunto convexo, del cual las ergódicas son los puntos extremales.

Cuando estudiamos la dinámica de T habiendo fijado una medida ergódica μ, se puede estimar la probabilidad $\mu(A)$ de un evento A observando la *frecuencia* con la que ocurre a lo largo de la órbita de un punto. Esta afirmación suele resumirse diciendo que, para una medida ergódica, *los promedios espaciales son iguales a los promedios temporales.* Este es el contenido de distintos teoremas ergódicos, que por el momento englobaremos en el siguiente enunciado informal:

Teorema (Teorema ergódico). *Sean $(X, \mathcal{A})$ un espacio medible, $T : X \to X$ una transformación medible y μ una medida de probabilidad T-invariante y ergódica. Sea $f : X \to \mathbb{R}$ una función. Entonces, para casi todo x con respecto de μ,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = \int_X f \, d\mu.$$

Aquí no estamos explicando qué tipo de función f se considera, ni a qué noción de convergencia se refiere el límite. Distintas maneras posibles de hacer preciso este enunciado dan lugar a distintos teoremas, que son importantes en distintos contextos y que se demuestran con técnicas muy distintas. Volveremos a ellos más adelante.

3. El shift en el espacio de símbolos

Sea $\mathcal{A}$ un conjunto finito que llamaremos *alfabeto*. Por simplicidad, a menos que se explicite lo contrario trabajaremos con $\mathcal{A} = \{0, 1\}$. Dotamos a $\mathcal{A}$ de la topología discreta. El *espacio de símbolos* es el conjunto

$$\Sigma = \mathcal{A}^{\mathbb{N}} = \{x = (x_i)_{i\in\mathbb{N}} : \ \forall i \in \mathbb{N} \ x_i \in \mathcal{A}\}$$

de las sucesiones de ceros y unos, con la topología producto. Es compacto, perfecto y totalmente disconexo; es decir, es un conjunto de Cantor.

Usaremos los nombres $x, y, z...$ para los puntos de Σ, y éstos se referirán a las suceisiones de término general $x_i, y_i, z_i...$ Usualmente escribiremos a x como

$$x = (x_0, x_1, x_2, \ldots)$$

Dados $N \in \mathbb{N}$ y $(a_0, a_1, \dots, a_N) \in \mathcal{A}^{N+1}$, definimos el *cilindro* $C[a_0, a_1, \dots, a_N]$ como

$$C[a_0, a_1, \dots, a_N] = \{x \in \Sigma : \; x_i = a_i \; \forall i = 0, \dots, N\}.$$

Los cilindros forman una base de la topología. La topología es metrizable, y una métrica que la induce está dada por

$$d(x, y) = \begin{cases} 0 & \text{si } x = y \\ \frac{1}{2^k} & \text{si } x \neq y \text{ y } k = \min\{j : \; x_j \neq y_j\} \end{cases}$$

Cuando $x_0 \neq y_0$, $d(x, y) = 1$ es máxima. Con esta distancia, el cilindro $C[a_0, a_1, \dots, a_N]$ es la bola de radio $1/2^{N+1}$ centrada en cualquiera de sus puntos.

Definición 3. El shift es la función $\sigma : \Sigma \to \Sigma$, dada por

$$\sigma(x) = y \iff \forall i \in \mathbb{N} \; y_i = x_{i+1}.$$

Esto es, $\sigma(x_0, x_1, x_2, x_3, \dots) = (x_1, x_2, x_3, \dots)$

Claramente, es una transformación continua en Σ que no es inyectiva pero sí sobreyectiva.

Observación 1. Si pensamos en $x \in \{0, 1\}^{\mathbb{N}}$ como siendo el número $x = 0, x_1 x_2 x_3 \dots \in [0, 1]$ escrito en base 2, el shift está dado por $\sigma(x) = 2x (mod 1)$.

Al estudiar la dinámica de σ observamos que se cumplen las propiedades siguientes:

1. *Los puntos periódicos son densos.*

   $x \in \Sigma$ es un punto periódico si $\sigma^n(x) = x$ para algún $n \in \mathbb{Z}$. Para ver que los puntos periódicos son densos, alcanza con ver que cualquier cilindro continene un punto periódico. Y, efectivamente, el cilindro $C[a_0, a_1, \dots, a_N]$ contiene al punto periódico

   $$(a_0, a_1 \dots, a_N, a_0, a_1 \dots, a_N, \dots)$$

2. *Hay puntos de órbita densa.*

   Para ver esto, enumeremos de alguna forma las sucesiones finitas de ceros y unos:

   0
   1
   00
   01
   10
   11
   000
   001
   011
   ⋮

   Cualquier sucesión $x \in \Sigma$ en la que aparezcan todas ellas tendrá órbita densa bajo la acción de σ. Un ejemplo es la sucesión

   $$(0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, \dots)$$

3. *Hay sensibilidad respecto de las condiciones iniciales.*

   Esta es la propiedad básica de lo que se considera una *dinámica caótica*, y es una condición que aparece en todas las definiciones topológicas del caos. Significa que cada punto de $x \in \Sigma$ es aproximado, tanto como se desee, por puntos que tienen una trayectoria "muy distinta" a la de x. Estando en un espacio métrico, pensaremos que dos puntos son "muy distintos" si están a una distancia grande. Podemos decir que hay sensibilidad respecto de las condiciones iniciales si

   $$\exists \varepsilon > 0 \ / \ \forall x \in \Sigma \ \forall \delta > 0 \ \exists y \in \Sigma \ \exists n \in \mathbb{N} \ / \ d(x,y) < \delta \text{ y } d(\sigma^n(x), \sigma^n(y)) < \varepsilon.$$

   En realidad el shift cumple algo mucho más fuerte:

   $$\forall x, y \in \Sigma \ \ (x \neq y \Rightarrow \exists n \in \mathbb{N} \ / \ d(\sigma^n(x), \sigma^n(y)) = 1).$$

   En efecto, alcanza con tomar n de modo que $x_n \neq y_n$.

Observación 2. Lo que hemos definido se conoce como shift unilateral. Es posible tomar $\Sigma = \mathcal{A}^{\mathbb{Z}}$, y definir del mismo modo a σ, como el corrimiento "hacia la izquierda". Es decir,

$$\sigma(x) = y \Longleftrightarrow \forall i \in \mathbb{Z} \ y_i = x_{i+1}.$$

Esta función se conoce como shift bilateral o *full shift*. En este caso σ es invertible, y es posible estudiar sus iterados positivos y negativos. Todo lo que haremos a continuación será presentado para el shift unilateral por comodidad, pero en realidad también vale para el shift bilateral.

3.1. El shift de Bernoulli

Procederemos ahora a definir distintas medidas invariantes y ergódicas para el shift $\sigma : \Sigma \to \Sigma$. Serán medidas en la σ-álgebra de Borel del espacio de símbolos. Bastará con definirlas apropiadamente en los cilindros, y tendrán una extensión única a todos los borelianos en virtud del teorema de extensión de medidas de Carathéodory.

Consideremos el experimento de arrojar repetidamente una moneda. Si sale cara anotamos un 0 y si sale cruz un 1. Las sucesivas tiradas determinan una sucesión de ceros y unos donde las distintas coordenadas son independientes. Si la probabilidad de que salga cara es p y la probabilidad de cruz es $1-p$, la probabilidad de que la sucesión comience con

$$0, 0, 1, 1, 1, 1, 0$$

será $p^3(1-p)^4$. Es natural entonces definir la medida de un cilindro $C[a_0, a_1, \ldots, a_N]$ como

$$\mu_p(C[a_0, a_1, \ldots, a_N]) = p^{N+1-\sum_{i=0}^N a_i}(1-p)^{\sum_{i=0}^N a_i}.$$

La medida que esto define en los borelianos de Σ, que llamaremos μ_p, se llama *medida de Bernoulli.*

Proposición 1. *Para $p \in (0,1)$ la medida μ_p es ergódica.*

Esta proposición admite una demostración elemental, que no haremos por falta de tiempo.

Observación 3. Al ser positiva en los cilindros, esta medida es positiva en abiertos y por lo tanto su soporte es todo Σ. Sin embargo, para distintos valores de p las medidas μ_p son mutuamente singulares. Esto quiere decir que si $p, p' \in (0,1)$, $p \neq p'$, hay conjuntos medibles disjuntos $A, B \subset \Sigma$ tales que $\mu_p(A) = \mu_{p'}(B) = 1$.

Para ver esto, consideremos la sucesión $\{X_i\}_{i\geq 0}$ de variables aleatorias que nos dicen si en la tirada i obtuvimos cara o cruz. Son independientes, idénticamente distribuidas, toman el valor 0 con probabilidad p y el valor 1 con probabilidad $1-p$. La ley fuerte de los grandes números nos dice que con probabilidad 1

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X_i = 1 - p.$$

Observemos que $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X_i$ es la proporción de veces que $X_i = 1$ para $i = 0, \ldots, n-1$. Por lo tanto, lo que está diciendo la ley de los grandes números es que para casi toda sucesión $x \in \Sigma$ respecto de la medida μ_p, la proporción de unos en $(x_0, \ldots, x_{n-1})$ tiende a $1-p$.

Si $p \neq p'$, $A = \{x \in \Sigma : \frac{1}{n} \sum_{i=0}^{n-1} x_i \to 1-p\}$ y $B = \{x \in \Sigma : \frac{1}{n} \sum_{i=0}^{n-1} x_i \to 1-p'\}$, entonces A y B son dos conjuntos medibles disjuntos tales que $\mu_p(A) = 1$ y $\mu_{p'}(B) = 1$.

Otra manera de contar la proporción de unos que tiene una sucesión es mirar la proporción de tiempo que su órbita pasa en el cilindro $C[1]$. Es decir, la cantidad

$$\frac{\sharp\{i \in \{0, 1, \ldots n-1\} : \ \sigma^i(x) \in C[1]\}}{n},$$

que convergerá a $1-p$ para casi todo x respecto de la medida μ_p. Vemos aquí una instancia en que se verifica el teorema ergódico: si f es la función característica de $C[1]$, lo que estamos diciendo es que

$$\frac{1}{n} \sum_{i=0}^{n-1} f(\sigma^i(x)) = 1 - p = \int_\Sigma f \, d\mu_p$$

para μ_p-casi todo $x \in \Sigma$.

Ejemplo 3. Consideremos $T : [0,1] \to [0,1]$ dada por $T(x) = 2x(\text{mod}1)$. Expresando a $x \in [0,1]$ como su desarrollo en base 2 $x = 0, x_0x_1x_2\ldots$, T coincide con el shift en todos los puntos cuyo desarrollo en base 2 es único. Estamos exceptuando sólo una cantidad numerable de puntos. Por lo tanto, desde el punto de vista de la teoría de la medida, podemos decir que estas dos transformaciones "son la misma" siempre que trabajemos con medidas que asignen medida cero a los puntos. Es el caso de todas las medidas μ_p con $p \in (0,1)$.

Por lo tanto, la transformación T admite una familia infinita a un parámetro real de medidas invariantes mutuamente singulares. La medida de Lebesgue es la que corresponde a $p = 1/2$.

Ejemplo 4 (Números normales.)**.** Tomemos $p = 1/2$, y pensemos nuevamente en una sucesión de ceros y unos como en un punto $x \in [0,1]$ escrito en base 2.

Con esta interpretación, $\mu_{\frac{1}{2}}$ es la medida de Lebesgue en $[0,1]$. Decimos que x es *normal en base 2* si la proporción de ceros en su desarrollo decimal es igual a la de unos, es decir, ambas son iguales a $1/2$. De la discusión anterior sabemos que

casi todo punto de $[0,1]$ *con respecto a la medida de Lebesgue es normal en base 2.*

Llamemos N_2 al conjunto de los números normales en base 2.

De forma análoga, decimos que $x \in [0,1]$ es *normal en base* k ($k \geq 2$) si todos los dígitos $0, 1, \ldots, k-1$ aparecen en igual proporción en su desarrollo decimal, es decir, con proporción $1/k$. Partiendo del alfabeto $\mathcal{A} = \{0, 1, \ldots k-1\}$ podemos definir en forma análoga una medida de Bernoulli en el espacio de símbolos donde cada coordenada de $x \in \Sigma$ toma cualquier valor de $\mathcal{A}$ con probabilidad $1/k$. En el intervalo $[0,1]$, identificado con Σ a través del desarrollo decimal en base k, esto da la medida de Lebesgue. Por lo tanto tendremos que

casi todo punto de $[0,1]$ *con respecto a la medida de Lebesgue es normal en base* k.

Llamemos N_k al conjunto de los números normales en base k.

Decimos que $x \in [0,1]$ es *normal* si es normal en base k para todo $k \geq 2$. Esto es, si pertenece al conjunto

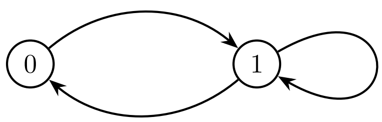
$$N = \cap_{k=2}^{\infty} N_k.$$

N tiene medida de Lebesgue total por ser intersección numerable de conjuntos de medida de Lebesgue total. O sea, casi todos los números son normales. ¡Sin embargo, no se conoce ningún número normal!

3.2. Subshifts de tipo finito y medidas de Markov

A continuación definiremos otra familia de medidas, con propiedades distintas a las de las medidas de Bernoulli. Para estas medidas, las coordenadas de un elemento del espacio de símbolos no serán independientes, sino que tendrán una dependencia markoviana.

Consideremos un grafo orientado G cuyos vértices sean los elementos de $\mathcal{A}$. Para fijar ideas, trabajaremos con $\mathcal{A} = \{0, 1\}$ y con el grafo



Miraremos caminatas al azar en este grafo, que comienzan en cualquiera de sus vértices y respetan las aristas. Una tal caminata es una sucesión de ceros y unos, que podría por ejemplo empezar así:

$$(1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1 \ldots)$$

Como no hay un lazo en el vértice 0, no puede haber en esta sucesión dos ceros seguidos. Las posibles caminatas en G son los elementos del conjunto

$$\Sigma_G = \{x \in \Sigma : \ \forall i \in \mathbb{N}\, (x_i = 0 \Rightarrow x_{i+1} \neq 0)\}.$$

La matriz de adyacencia del grafo G es

$$A = A_G = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Observación 4. Trabajando con cualquier alfabeto finito $\mathcal{A}$, en el que se fija un orden, y cualquier grafo orientado G cuyo conjunto de vértices es $\mathcal{A}$, es posible definir la matriz de adyacencia. Es una matriz de ceros y unos cuya entrada ij es 1 si hay una arista del i-ésimo vértice al j-ésimo vértice y es 0 en caso contrario. Tiene la misma información que el grafo G, y por lo tanto determina el conjunto Σ_G de sucesiones $x \in \Sigma$ que pueden aparecer como caminatas en el grafo. Σ_G es un cerrado invariante bajo el shift, cuyos elementos se llaman sucesiones admisibles.

Definición 4. σ_G es la restricción de σ al conjunto Σ_G. Cuando $\Sigma_G \neq \Sigma$, decimos que $\sigma_G : \Sigma_G \to \Sigma_G$ es un subshift de tipo finito.

El único caso en que $\Sigma_G = \Sigma$ es aquel en que G es un grafo orientado completo.

Las propiedades dinámicas de σ_G dependen, por supuesto, del grafo G. Un ejemplo de esto es la siguiente proposición:

Proposición 2. *σ_G tiene una órbita densa $\iff$ G es fuertemente conexo, es decir, si dados dos vértices v y w de G hay un camino en G que va de v a w.*

Las medidas de Bernoulli no "ven" los subshifts de tipo finito, es decir,

Proposición 3. *Si μ es una medida de Bernoulli y $\Sigma_G \neq \Sigma$, entonces $\mu(\Sigma_G) = 0$.*

Para probar esto, recordemos el clásico Lema de Borel-Cantelli. Dice que:

Lema (Borel-Cantelli)**.** *En un espacio de probabilidad (X, μ), sea $\{A_n\}_{n\geq 0}$ una sucesión de eventos independientes. Si $\sum_{n=0}^{\infty} \mu(A_n) = \infty$, entonces*

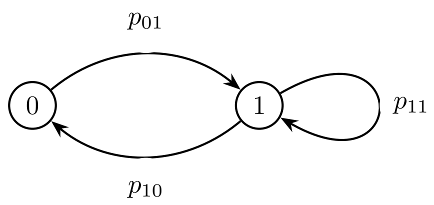
$$\mu\left(\cap_{n=0}^{\infty} \cup_{k\geq n} A_k\right) = 1.$$

Demostración de la Proposición. Demostremos la proposición para el grafo G de nuestro ejemplo. Digamos que μ es la medida de Bernoulli μ_p. Tomemos, en el espacio de probabilidad (Σ, μ), los conjuntos $A_n = \sigma_{2n}(C[0,0])$ para $n \geq 0$. Es decir, A_n es el conjunto de sucesiones que tienen ceros en los lugares $2n$ y $2n+1$, por lo que los elementos de A_n no son sucesiones admisibles. Para todo n, $\mu(A_n) = p^2$, por lo que claramente $\sum_{n=0}^{\infty} \mu(A_n) = \infty$.

$x \in \cap_{n=0}^{\infty} \cup_{k\geq n} A_k$ si y sólo si para todo n existe un $k \geq n$ tal que $x \in A_k$. En particular, x no es una sucesión admisible. Por el lema de Borel-Cantelli, hay un conjunto de medida total de sucesiones no admisibles, por lo que el conjunto de sucesiones admisibles tiene medida cero.

Por lo tanto, para estudiar los subshifts de tipo finito desde el punto de vista ergódico, será necesario considerar otras medidas invariantes.

Agregaremos pesos positivios a las aristas del grafo G de modo que, para cualquier vértice v, la suma de los pesos de las aristas que salen de v sea 1.



Como hay una única arista que sale del vértice 0, necesariamente $p_{01} = 1$. La otra condición es que $p_{10} + p_{11} = 1$.

Fijemos $\pi_0 \in (0, 1)$, que interpretaremos como la probabilidad de que un camino en G comience en el vértice 0, y sea $\pi_1 = 1 - \pi_0$. Interpretaremos p_{vw} como la probabilidad de, estando el el vértice v, ir al vértice w en el siguiente paso. No depende de *cómo* llegamos a v. (La caminata al azar en G es una *cadena de Markov homogénea*). Si no hay una arista de v a w, escribiremos $p_{vw} = 0$. En nuestro caso, $p_{00} = 0$.

La probabilidad de que una caminata al azar en G comience con

$$0, 1, 1, 0, 1, 0, 1, 1$$

será $\pi_0 \cdot p_{01} \cdot p_{11} \cdot p_{10} \cdot p_{01} \cdot p_{10} \cdot p_{01} \cdot p_{11}$. Es natural entonces definir la medida de un cilindro $C[a_0, a_1, \ldots, a_{N-1}]$ como

$$\mu_p(C[a_0, a_1, \ldots, a_N]) = \pi_{a_0} p_{a_0 a_1} \ldots p_{a_{N-1} a_N}.$$

La medida que esto define en los borelianos de Σ depende del vector $\pi = (\pi_0, \pi_1)$ y de la matriz $P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$, llamada *matriz de transición*. Diremos que π es *estacionario* si $\pi P = \pi$. Observemos que la matriz P es una *matriz estocástica*, es decir, sus entradas son no negativas y sus filas suman 1. Por lo tanto tiene valor propio 1, lo cual garantiza que hay vectores estacionarios no triviales.

Cuando G es fuertemente conexo la matriz P es irreducible, y el teorema de Perron-Frobenius garantiza la existencia de un vector estacionario π de entradas estrictamente positivas, único a menos de multiplicación por un escalar positivo. Puede normalizarse para que sus entradas sumen 1. Con esta elección de π, la medida μ en Σ depende únicamente de la matriz de transición P. Esta medida se llama *medida de Markov*.

Observación 5. Para la medida de Markov μ, $\mu(\Sigma_G) = 1$. Por lo tanto, podemos pensar que es una medida en Σ_G.

Proposición 4. *Sean G un grafo fuertemente conexo cuyo conjunto de vértices es el alfabeto finito $\mathcal{A}$, y $A = (a_{ij})$ su matriz de adyacencia. Sean $P = (p_{ij})$ una matriz estocástica tal que $p_{ij} = 0 \Leftrightarrow a_{ij} = 0$ y π el vector estacionario de P cuyas entradas suman 1. Entonces la medida de Markov μ en Σ determinada por P y π es shift-invariante y ergódica.*

Es posible definir muchas otras medidas shift-invariantes y ergódicas.

El shift es un sistema dinámico interesante en sí mismo, pero se usa también como modelo combinatorio de muchos sistemas dinámicos de origen geométrico. Qué medidas son relevantes, o significativas, depende de la dinámica que se quiera estudiar.

4. Entropía métrica: una medida del caos

El término *entropía* viene de la física, más concretamente de la termodinámica. Fue introducido por Rudolf Clausius a mediados del siglo XIX, y a grandes razgos se refiere a la energía que no es capaz de producir trabajo. Es central en la segunda ley de la termodinámica.

A mediados del siglo XX, el concepto de entropía fue retomado y resignificado en el campo de la ingeniería eléctica por Claude Shannon, que fue un pionero de la teoría de la información. En la *entropía de Shannon* se inspiró Kolmogórov para definir la entropía métrica de una transformación, que es la noción que se presenta aquí. La entropía pretende ser, desde sus orígenes en la termodinámica, una medida del *desorden* de un sistema.

4.1. Cantidad de información

Hay varias definiciones posibles de sistema dinámico *caótico*, pero en un espacio métrico siempre es una propiedad central la sensibilidad respecto de las condiciones iniciales. Ésta nos dice que es muy difícil *predecir* la evolución del sistema a partir de información parcial sobre su estado actual (por ejemplo, a partir de una medición inexacta del mismo).

Desde el punto de vista de la teoría ergódica, quisiéramos dar una medida del caos de un sistema en términos probabilísticos. Para esto, seguiremos estudiando el shift en el espacio de símbolos, pero sólo para simplificar la presentación.

Pensaremos en $x \in \Sigma$ como en un mensaje, que está siendo transmitido a través de algún canal de comunicación. Inicialmente hemos leído la letra x_0, y a medida que pasa el tiempo nos van llegando las siguientes letras, y en tiempo n hemos recibido el mensaje $x_0 x_1 \dots x_n$. ¿Qué *información* hemos recibido? La idea de Shannon es que no todos los mensajes tienen la misma información, y que la cantidad de información de un mensaje es suceptible de ser cuantificada.

Pongamos un ejemplo: pensemos en un mensaje que comienza por

La mayoría de los encuestados opina que la obra cumbre de la literatura en lengua castellana es

¿Cuánta información nos aportarán las siguientes dos palabras? Si las siguientes dos palabras son

El Quijote,

la verdad es que no mucha. No nos habremos enterado de casi nada. Cualquier otra respuesta nos sorprendería, sería una noticia, por lo que aportaría mucha más información. La información es mayor cuanto más improbable sea el contenido del mensaje.

Shannon pensó que la cantidad de información de un evento E es en un espacio de probabilidad $(\Sigma, \mathcal{B}, \mu)$, o equivalentemente la cantidad de información de la afirmación "$x \in E$", debe definirse por medio de una función $I : \mathcal{B} \to [0, +\infty)$ que cumpla las siguientes propiedades:

1. $I(E)$ es una función decreciente de $\mu(E)$ que vale 0 cuando $\mu(E) = 1$.
2. Si A y B son eventos independientes, $I(A \cap B) = I(A) + I(B)$.

A partir de esto es sencillo ver que $I(E) = -k \log(\mu(E))$. En teoría ergódica se suele tomar $k = 1$, y en teoría de la información $k = 1/\log 2$, lo que corresponde a tomar el logaritmo en base 2.

Definición 5. En un espacio de probabilidad $(\Sigma, \mathcal{B}, \mu)$ se define la cantidad de información como $I : \mathcal{B} \to [0, +\infty)$ dada por

$$I(E) = -\log(\mu(E)).$$

Ejemplo 5. Si $\Sigma = \{0,1\}^{\mathbb{N}}$ con su σ-álgebra de Borel $\mathcal{B}$ y μ es la medida de Bernoulli $\mu = \mu_{1/2}$, la cantidad de información de un cilindro $A = C[x_0, \dots, x_N]$ es $I(A) = -\log(1/2^N) = (N+1)\log 2$. ($I(E)/\log 2$ se mide en *bits*, por lo que la cantidad de información de A es $N+1$ bits).

4.2. Entropía del shift

En esta sección $(\Sigma, \mathcal{B})$ será el espacio de símbolos con su σ-álgebra de Borel, y $\sigma : \Sigma \to \Sigma$ será el shift. Daremos el concepto de *entropía métrica* de una transformación en este caso, pero únicamente con el fin de simplificar la presentación. Este concepto fue definido por Kolmogórov con toda generalidad de modo similar.

Definición 6. Una partición (finita) $\mathcal{P}$ de Σ es una familia finita de conjuntos medibles disjuntos cuya unión es Σ.

1. Si $\mathcal{P}$ y $\mathcal{Q}$ son particiones, definimos su producto como la partición $\mathcal{P} \vee \mathcal{Q}$ cuyos elementos son los conjuntos de la forma $A \cap B$, con $A \in \mathcal{P}$ y $B \in \mathcal{Q}$.
2. $\sigma^{-1}\mathcal{P}$ es la partición cuyos elementos son los conjuntos de la forma $\sigma^{-1}(A)$ con $A \in \mathcal{P}$.

Fijemos una medida de probabilidad μ en Σ. Ahora que hemos definido la cantidad de información de un conjunto, podemos preguntarnos, dada una partición $\mathcal{P}$, qué cantidad de información esperamos tener al enterarnos de a qué elemento de $\mathcal{P}$ pertenece $x \in \Sigma$. Esto es la entropía de $\mathcal{P}$.

Definición 7. La entropía de una partición $\mathcal{P} = \{A_1, \dots, A_k\}$ es el promedio ponderado de la cantidad de información de los elementos de $\mathcal{P}$ dado por[1]

$$H(\mathcal{P}, \mu) = -\sum_{i=1}^{k} \mu(A_i) \log(\mu(A_i)).$$

Es lo mismo que el valor esperado de la variable aleatoria

$$\begin{array}{ccc} \Sigma & \to & \mathbb{R} \\ x & \mapsto & I(A_x) \end{array}$$

si A_x es el elemento de $\mathcal{P}$ que contiene a x.

Esta definición cumple las propiedades siguientes:

Proposición 5. *1. $H(\mathcal{P}, \mu) \leq 0$, y la igualdad se da si y sólo si hay un único A_i con medida positiva, y en ese caso $\mu(A_i) = 1$.*

[1] Adoptamos la convención de que $0 \cdot \log(0) = 0$.

2. *$H(\mathcal{P},\mu) \leq \log k$, y la igualdad se da si y sólo si $\mu(A_i) = 1/k$ para todo i.*

3. *$H(\mathcal{P},\mu) \leq H(\mathcal{P} \vee \mathcal{Q},\mu) \leq H(\mathcal{P},\mu) + H(\mathcal{Q},\mu)$.*

4. *Si μ es una medida shift-invariante, $H(\mathcal{P},\mu) = H(\sigma^{-n}\mathcal{P},\mu)$ para todo $n \geq 0$.*

El cuarto inciso es inmediato a partir de la definición de H. Los otros tres se prueban a partir de la desigualdad de Jensen aplicada a la función cóncava $\varphi(x) = -x\log(x)$.

Observemos que el tercer inciso tiene la consecuencia siguiente, que es bastante intuitiva: la cantidad de información esperada que proporciona $\mathcal{P} \vee \mathcal{Q}$ es mayor que la que proporciona $\mathcal{P}$.

Consideremos en el espacio de símbolos la partición $\mathcal{P} = \{C[0], C[1]\}$ de Σ. $x \in C[i]$ nos dice que $x_0 = i$. Para $n > 0$, definimos

$$\mathcal{P}_n = \mathcal{P} \vee \sigma^{-1}\mathcal{P} \vee \cdots \vee \sigma^{-(n-1)}\mathcal{P}.$$

Los elementos de $\mathcal{P}_n$ son los cilindros de largo n.

La sucesión

$$H(\mathcal{P}_n,\mu)$$

crece con n. Además, de la Proposición 5 se desprende queremos

$$H(\mathcal{P}_n,\mu) \leq nH(\mathcal{P},\mu).$$

En particular $H(\mathcal{P}_n,\mu)/n$ está acotada por $H(\mathcal{P},\mu)$.

También a partir de la Proposición 5 es sencillo probar que la sucesión $H(\mathcal{P}_n,\mu)$ es *subaditiva*, es decir,que

$$H(\mathcal{P}_{n+m},\mu) \leq H(\mathcal{P}_n,\mu) + H(\mathcal{P}_m,\mu) \quad \forall n,m \geq 1.$$

Una sucesión subaditiva $\{a_n\}$ cumple siempre que

$$\exists \lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\geq 1} \frac{a_n}{n}.$$

Por lo tanto tiene sentido la siguiente definición:

Definición 8. La entropía del shift respecto a la partición $\mathcal{P}$ y la medida μ es

$$h(\sigma,\mathcal{P},\mu) = \lim_{n\to\infty} \frac{H(\mathcal{P}_n,\mu)}{n}.$$

Ejemplo 6. Calculemos la entropía del shift para $\mathcal{P} = \{C[0], C[1]\}$ como arriba y para la medida de Bernoulli $\mu = \mu_{\frac{1}{2}}$. La medida de un elemento de $\mathcal{P}_n$ es $1/2^n$, y su cantidad de información es $n\log 2$. Como la partición $\mathcal{P}_n$ tiene 2^n elementos, $H(\mathcal{P}_n,\mu) = n\log 2$ y $h(\sigma,\mathcal{P},\mu) = \log 2$.

De entre todas las medidas de Bernoulli μ_p, $\mu_{\frac{1}{2}}$ es la que maximiza la entropía de la partición $\mathcal{P}_n$, y por lo tanto la que maximiza $h(\sigma,\mathcal{P},\mu)$.

¿Qué quiere decir que $h(\sigma,\mathcal{P},\mu) > 0$?

Volvamos a la interpretación según la cual $x \in \Sigma$ es un mensaje, que se está transmitiendo a través de un canal. En el instante n nos llega x_{n-1}, es decir, en el instante n sabemos a qué elemento de $\mathcal{P}$ pertenece $\sigma^{n-1}(x)$. Si

$h(\sigma, \mathcal{P}, \mu) > 0$, esto dice que la cantidad esperada de información del mensaje crece linealmente con el tiempo. Es decir, no podemos predecir el contenido del mensaje a partir de una porción finita del mismo. Más aun, la cantidad de información que nos proporciona el mensaje nos sigue llegando a un ritmo aproximadamente constante.

Consideremos otra interpretación: pensemos en un sistema que tiene dos estados posibles que son 0 y 1. En tiempo n el sistema está en estado i si $\sigma^{n-1}(x) \in C[i]$. Si $h(\sigma, \mathcal{P}, \mu) > 0$, esto dice que no podemos predecir un estado futuro del sistema a partir de lo que hemos observado hasta el momento. Es decir, no es posible predecir x_{n+m} a partir de $x_0, x_1, \ldots, x_n$. Además, haber observado el sistema hasta tiempo n, aunque n sea grande, no nos pone en mejores condiciones para predecir su estado en tiempo $n + m$, pues la velocidad a la que aumenta la información es positiva. Es en este sentido que la entropía es una medida del caos.

Por supuesto que podríamos haber definido $h(\sigma, \mathcal{P}, \mu)$ a partir de otra partición $\mathcal{P}$. Fijada una medida μ, la entropía métrica de σ con respecto a μ se define de la siguiente manera:

Definición 9. La entropía métrica de σ con respecto a μ es

$$h(\sigma, \mu) = sup_{\mathcal{P}} h(\sigma, \mathcal{P}, \mu),$$

donde el supremo se toma sobre todas las particiones finitas de Σ por conjuntos medibles.

Es interesante señalar que todas las medidas que hemos introducido en el espacio de símbolos (las medidas de Bernoulli y las medidas de Markov) tienen entropía positiva.

Un comentario final: hay también una noción topológica de entropía de una transformación continua en un espacio métrico compacto, que da una medida del caos de un sistema cuando se lo estudia desde el punto de vista de la dinámica topológica. No la damos aquí porque es un poco engorrosa. Sin embargo, la entropía topológica y la entropía métrica están relacionadas mediante el siguiente *principio variacional de la entropía*:

$$h_{top}(\sigma) = \sup_{\mu} h(\sigma, \mu),$$

donde el supremo se toma sobre todas las medidas de probabilidad σ-invariantes.

5. Teoremas ergódicos

En esta sección, $(X, \mathcal{A})$ será un espacio medible, y $T : X \to X$ será una transformación medible que preserva una medida de probabiliad μ.

Un teorema ergódico es una teorema que describe el comportamiento de una sucesión

$$\frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j,$$

donde $f : X \to \mathbb{R}$ o $f : X \to \mathbb{C}$ es una función. Su formulación precisa depende de

* qué clase de función es f (continua, medible, de $L^p(\mu)$...)

* qué noción de convergencia se considera (puntual, uniforme, en L^p...)

Para $x \in X$, $\frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j(x)$ es el *promedio temporal* de f a lo largo de la órbita de x hasta tiempo $n-1$. Interesa estudiar su convergencia y su dependencia de x. Como ya dijimos en la Sección 2, si μ es ergódica este límite será igual en casi todo punto, e igual a $\int_X f\, d\mu$.

Observemos que cuando existe el límite, es decir, cuando está definida la función

$$\tilde{f}(x) = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j(x)$$

en algún espacio funcional, $\tilde{f} : X \to \mathbb{R}$ es constante en las órbitas de T (es decir, es invariante por T). Al ser límite de funciones medibles es una función medible, por lo que si μ es ergódica debe tomar sólo un valor. Pero cualquier función constante en un espacio de probabilidad será igual a su integral. Por lo tanto, para concluir que $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j(x) = \int f\, d\mu$, lo esencial es probar la existencia de $\tilde{f}$, en algún contexto.

Dos teoremas ergódicos importantes de la década de 1930 son los teoremas ergódicos de Birkhoff y de von Neumann, de los que hablaremos a continuación.

5.1. Teorema de Birkhoff

Este teorema lleva el nombre de George David Birkhoff, aunque a veces también se lo llama Teorema de Birkhoff–Khinchin, por el matemático soviético Aleksandr Khinchin.

Teorema Ergódico de Birkhoff. *Sea $(X, \mathcal{A}, \mu)$ un espacio de probabilidad, y $T : X \to X$ una transformación que preserva la medida μ. Si $f : X \to \mathbb{R}$ es una función integrable, para μ-casi todo $x \in X$ existe el límite*

$$\tilde{f}(x) = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j(x).$$

La función $\tilde{f} : X \to \mathbb{R}$ es medible y constante en las órbitas de T. En particular, si μ es ergódica para T,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j(x) = \int_X f\, d\mu$$

para μ-casi todo $x \in X$.

Un caso particular de este teorema que es muy interesante es aquel en que f es la función característica de algún conjunto medible A. En este caso, $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j(x)$ es la proporción de tiempo que la órbita de x pasa en A. El teorema dice que cuando μ es ergódica esto coincide con $\mu(A)$, y en este sentido μ nos dice cómo es *probable* que se comporte el sistema a tiempos largos.

Probar este teorema no es fácil y por razones de tiempo está fuera del alcance de este minicurso. Daremos la demostración de un teorema de enunciado similar pero de demostración muy diferente: el Teorema Ergódico de von Neumann.

5.2. Teorema Ergódico de von Neumann

Por simplicidad, en esta sección asumiremos que $T : X \to X$, además de preservar la medida de probabilidad μ, es biyectiva.

El teorema que veremos a continuación tiene un enunciado muy similar al anterior, sólo que trata de funciones de cuadrado integrable y convergencia en el espacio $L^2(\mu)$. Recordemos que

$$\mathcal{L}^2(\mu) = \{f : X \to \mathbb{C} : \int_X |f|^2 \, d\mu < \infty\},$$

y $L^2(\mu)$ es el cociente de este espacio por la relación de igualdad en μ-casi todo punto. Está dotado del producto interno

$$\langle f, g \rangle = \int_X f\bar{g} \, d\mu,$$

con el cual es completo. Es decir, es un espacio de Hilbert.

Consideremos el subespacio cerrado de las funciones invariantes bajo T

$$I = \{f \in L^2(\mu) : \; f \circ T = f\}.$$

La proyección ortogonal sobre este subespacio es $P_I : L^2(\mu) \to L^2(\mu)$.

Observación 6. La medida μ es ergódica si y sólo si los únicos elementos de I son las funciones constantes.

El teorema ergódico de von Neumann se refiere a grupos de operadores unitarios en espacios de Hilbert, aunque enunciaremos aquí sólo el caso particular del mismo que es relevante en nuestro contexto. Este enunciado más restringido es también consecuencia de la demostración de Birkhoff de su teorema ergódico. Sin embargo, el teorema de von Neumann en su versión más general no se deduce del de Birkhoff. Por ser ésta más corta y elegante, daremos la prueba de von Neumann.

Teorema Ergódico de von Neumann. *Sea $(X, \mathcal{A}, \mu)$ un espacio de probabilidad, y $T : X \to X$ una transformación invertible que preserva la medida μ. Si $f \in L^2(\mu)$, existe el límite*

$$\tilde{f}(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j$$

en $L^2(\mu)$ y es igual a $P_I(f)$.

Observemos que para un espacio de probabilidad $L^2(\mu) \subset L^1(\mu)$. Es decir, todas las funciones que estamos considerando son integrables.

Observación 7. $\int_X f d\mu = \langle f, 1 \rangle = \langle P_I(f) + P_{I^\perp}(f), 1 \rangle = \langle P_I(f), 1 \rangle + \langle P_{I^\perp}(f), 1 \rangle$. Aquí $P_{I^\perp}(f)$ denota la proyección de f sobre el complemento ortogonal de I. Como éste es ortogonal a la función 1, tenemos que

$$\int_X f \, d\mu = \int_X P_I(f) \, d\mu.$$

Consideremos el operador lineal

$$\begin{array}{l} U_T : L^2(\mu) \to L^2(\mu) \\ U_T(f) = f \circ T. \end{array}$$

Tiene las siguientes propiedades, que se verifican con facilidad:

- U_T tiene autovalor 1, y $\ker(U_T - id) = I$.
- Al ser T invertible, U_T también es invertible, y además preserva el producto interno. Por lo tanto, U_T es un operador *unitario*, es decir, $U_T^* = (U_T)^{-1}$.
- Como $||U_T(f)|| = ||f||$ para toda f, todos los autovalores de U_T tienen norma 1.

Demostración del Teorema:
Queremos probar que

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} U_T^j(f) \longrightarrow_{n\to\infty} P_I(f).$$

Cuando $f \in I$, esto es evidente.

Probémoslo ahora para una función f de la forma $f = U_T(g) - g$. Como la suma $\sum_{j=0}^{n-1} U_T^j(f)$ es telescópica,

$$\left|\left| \frac{1}{n} \sum_{j=0}^{n-1} U_T^j(f) \right|\right| = \frac{1}{n} ||U_T^{n-1}(g) - g|| \leq \frac{2}{n} ||f|| \to 0.$$

Sea B la clausura de $\{U_T(g) - g \,:\, g \in L^2(\mu)\}$. Entonces para toda $f \in B$ se cumple que $\frac{1}{n} \sum_{j=0}^{n-1} U_T^j(f) \to 0$.

Vamos a ver que $B = I^\perp$, es decir, que $L^2(\mu) = I \oplus B$ donde la suma directa es ortogonal.

$$\begin{array}{rcl} f \perp B & \iff & \forall g \in L^2(\mu) \;\; \langle f, U_T(g) - g \rangle = 0 \\ & \iff & \forall g \in L^2(\mu) \;\; \langle f, U_T(g) \rangle = \langle f, g \rangle \\ & \iff & U_T(f) = f \\ & \iff & f \in I. \end{array}$$

Entonces

$$\begin{array}{rcl} \displaystyle \frac{1}{n} \sum_{j=0}^{n-1} U_T^j(f) & = & \displaystyle \frac{1}{n} \sum_{j=0}^{n-1} U_T^j(P_I(f)) + \frac{1}{n} \sum_{j=0}^{n-1} U_T^j(P_{I^\perp}(f)) \\ & = & \displaystyle P_I(f) + \frac{1}{n} \sum_{j=0}^{n-1} U_T^j(P_{I^\perp}(f)) \longrightarrow P_I(f). \end{array}$$

La idea de pasar de $T : X \to X$ a $U_T : L^2(m) \to L^2(m)$ es sumamente fructífera, yendo mucho más allá de la demostración de este teorema. Aunque no ahondaremos en este tema, daremos un pequeño ejemplo de su utilidad, probando que una rotación de ángulo irracional en la circunferencia es ergódica para la medida de Lebesgue.

Tomemos $\mathbb{S}^1 = \{z \in \mathbb{C} : \ |z| = 1\}$ y sea m la medida de Lebesgue en $\mathbb{S}^1$. Es decir, para $f : \mathbb{S}^1 \to \mathbb{R}$ integrable

$$\int_{\mathbb{S}^1} f(z)\, dm(z) = \int_0^1 f(e^{2\pi i x})\, dx.$$

La base de Fourier en $L^2(m)$ es $\{\varphi_n\}_{n\in\mathbb{Z}}$ dada por

$$\varphi_n(z) = z^n$$

o equivalentemente

$$\varphi_n(x) = e^{2\pi i n x}.$$

Es una base ortonormal. Por lo tanto, cualquier función $f \in L^2(m)$ se escribe como $f = \sum_{n\in\mathbb{Z}} c_n(f)\varphi_n$, donde $c_n(f) = \langle f, \varphi_n\rangle$.

Proposición 6. *Fijado $\alpha \in \mathbb{R}\backslash\mathbb{Q}$, sea $T : \mathcal{S}^1 \to \mathcal{S}^1$ la rotación de ángulo $2\pi\alpha$, dada por*

$$T(e^{2\pi i x}) = e^{2\pi i (x+\alpha)}.$$

Entonces m es ergódica para T.

Demostración: Claramente m es T-invariante. Para probar que m es ergódica, probaremos que el subespacio I de las funciones T-invariantes en $L^2(m)$ sólo contiene a las funciones constantes.

Tomemos $f \in I$, es decir, $f \in L^2(m)$ tal que $U_T(f) = f$. Queremos ver que f es constante en casi todo punto, es decir, que $c_n(f) = 0$ para $n \neq 0$.

$$\begin{aligned}
\langle f, \varphi_n\rangle &= \langle U_T(f), \varphi_n\rangle \\
&= \int_0^1 f(e^{2\pi i(x+\alpha)})e^{-2\pi i n x}\, dx \\
&= \int_\alpha^{\alpha+1} f(e^{2\pi i y})e^{-2\pi i n(y-\alpha)}\, dx \\
&= \int_0^1 f(e^{2\pi i y})e^{-2\pi i n(y-\alpha)}\, dx \\
&= e^{2\pi i n\alpha}\langle f, \varphi_n\rangle.
\end{aligned}$$

Esto sólo es posible si $\langle f, \varphi_n\rangle = 0 \ \ \forall n \neq 0$, que es lo que queríamos probar.